US ARMY
LABORATORY COMMAND

TECHNOLOGY TO WIN

CONTRACT REPORT NUMBER 35-89
PREPARED FOR THE

HUMAN ENGINEERING LABORATORY

BY

University of Massachusetts, Amherst, Massachusetts

IMPACT OF UNCERTAINTY AND DIAGNOSTICITY ON CLASSIFICATION OF MULTIDIMENSIONAL DATA WITH INTEGRAL AND SEPARABLE DISPLAYS OF SYSTEM STATUS

1989

Bruce G. Coury
Margery D. Boulette
Robert A. Smith

Contract DAAA15-85-K-0010

US ARMY HUMAN ENGINEERING LABORATORY
ABERDEEN PROVING GROUND, MARYLAND
HEL

Approved for public release;
distribution is unlimited.

ABERDEEN PROVING GROUND, MARYLAND 21005-5001

Impact of Uncertainty and Diagnosticity on Classification of Multidimensional Data With Integral and Separable Displays of System Status

by

Bruce G. Coury, Margery D. Boulette
and Robert A. Smith

To appear in Human Factors

Human Performance Laboratory
Technical Report No. BGC1-1989
University of Massachusetts
114 Marston Hall
Amherst, MA 01003
Phone: (413) 545-2854

ABSTRACT

Integrative, object-like displays have been advocated for presenting multidimensional system data. In this research, two experiments assessed the relative merits of integral and separable displays to enhance information processing ability when the identity of an instance of system data is uncertain. In each experiment, thirty subjects, equally divided into three groups, were trained to classify instances of system state into one of four state categories using a Configural display, a Bargraph display, or a Digital display. In Experiment 1, the distribution of instances from the range of possibilities within a state category were uniform; in Experiment 2, the distribution was biased toward those instances of highly uncertain state category membership. After training, subjects received extended practice classifying instances. In both experiments, uncertainty was found to have the greatest impact on the time to classify an instance of system data. In Experiment 1, the Bargraph display was consistently superior under all conditions of uncertainty. The Configural display was found to be superior to the Digital display under conditions of low uncertainty, while the Digital display was superior to the Configural display under conditions of high uncertainty. In Experiment 2, the superiority of the Bargraph display diminished, producing results equivalent to those of the Digital display; performance with the Configural display was worse than either of the other two displays. The impact of uncertainty on classification performance is discussed, especially in terms of mapping of display elements to system-state categories.

This research was supported by U.S. Army Human Engineering Laboratory Contract No. DAAA15-85-K-0010 to the first author. Requests for reprints should be sent to Bruce G. Coury, Dept. of Industrial Engineering and Operations Research, University of Massachusetts, 114 Marston Hall, Amherst, MA 01003.

INTRODUCTION

An operator's task, whether it be simple or complex, is comprised of two fundamental components (Rasmussen, 1986, 1988): 1) identifying the state of the system; and 2) selecting the appropriate course of action. Even as advanced technology and automated control systems provide the capability to integrate and summarize data, the operator must still understand the relationship between displayed values of system variables and the behavior of the system (Rasmussen, 1986, 1988; de Kleer and Brown, 1983). In fact, the introduction of automated control and decision aiding systems has increased, rather than decreased, the need for effective display design (Nickerson, 1986; Norman and Draper, 1986; Shneiderman, 1987; Hendler, 1988).

Consequently, one of the major challenges facing designers of operator interfaces for complex systems is that of selecting the appropriate display representation of system-status information. The research reported in this paper focuses on identification of system state, and examines the conditions under which a particular type of display enhances the accurate and timely assessment of the status of a system.

A Model of Operator Decision Making

Selecting the best display format for identification of system state can be aided by a model of operator performance that establishes a framework for assessing task demands. The model of the operator used in this research is based upon a number of approaches to operator performance (Miller, 1985; Rasmussen, 1986, 1988)) and user interaction with computer systems (Norman and Draper, 1986; Olson, 1987), as well as current theories of inductive reasoning (Holland, Holyoke, Nisbett and Thagard, 1986). Fundamental to those approaches and the model employed in this research is the assumption that identification of system state is equivalent to the determination of the current state of a decision problem, and is the prerequisite step to selecting the best course of action.

Determination of the state of a decision problem has been a critical and extensively researched component of effective decision making performance (Green and Swets, 1966; Holland et al., 1986) and a basic component of all types of decision aids (Sprague and Carlson, 1982; Miller, 1985; Henneman, 1988). Consequently, the operator's ability to map the values of critical decision variables to known definitions of decision states is the foundation for effective operator decision making performance, especially when system data must be integrated from a number of information sources, (Holland et al., 1986; Rasmussen, 1986, 1988).

Accurate identification of system state presupposes, however, that the the operator possess a well developed internal model of the criteria defining system-state categories. Thus, our model of operator performance views identification of system state as a categorization process, an approach consistent with current theories of human reasoning (Holland et al., 1986), concept learning (Smith and Medin, 1981), and the processing of information and structure (Garner, 1974, 1980; Pomerantz, Pristach and Carson, 1987). In our model of operator decision making, a system-state category is defined by a set of weighted attributes that serve as decision variables. An instance of a decision state is a specific set of values of decision variables from a particular state category. Operator performance is dependent upon an ability to match the data from each information source to an internal model defining state category membership.

In relatively simple operator tasks, the mapping of values to system-state categories is straightforward. Decision making becomes complex when that mapping is not straightforward and the correct decision is not immediately apparent. In such decision making environments, uncertainty arises because the values of system variables (be they single, direct readings of gauges or integrated measures of subsystem function) do not uniquely define a specific system-state.

Thus, uncertainty is a function of the degree of correspondence between the displayed values of current system status and the values of the attributes that define a specific system-state category. The degree of correspondence defines the diagnosticity of system variables. Combinations of values that uniquely define a state category are said to have high diagnostic value; conversely, data that do not uniquely define a state category are low in diagnostic value. In this research, then, diagnosticity and uncertainty are inversely related: combinations of values of system variables that are low in diagnosticity create conditions of high uncertainty for the operator.

There is considerable evidence to suggest, however, that the operator's ability to effectively monitor and diagnosis failures in a system is dependent upon the way in which information is displayed (Barnett and Wickens, 1988; Carswell and Wickens, 1987; Coury, Boulette, Zubritzky and Fisher, 1986; Boulette, Coury and Bezar, 1987; Woods, Wise and Hanes, 1981). For instance, in a recent study, Coury and Pietras (1989) found that the information necessary for supervising a simulated fluid processing plant operating in a normal mode was significantly different from the information required to control a failing process plant. Furthermore, the way in which information and data were presented to operators had a significant impact on their ability to effectively control the system and accommodate the disruptive effect of failures. Of particular interest was the fact that operators significantly altered their information gathering strategies when data and information were displayed in suboptimal form.

Integral and Separable Displays

There have been many recent studies that have addressed the relative merits of different types of display formats. The results of many of those studies support the use of integrative, object-like displays as a means for enhancing operator performance (Carswell and Wickens, 1987; Coury et al., 1986; Woods, Wise and Hanes, 1981). Such displays appear to be especially effective in situations where the decision problem faced by the operator is multidimensional and the values of the decision variables are correlated (Jacob, Egeth, and Bevan, 1976; Goldsmith and Schvaneveldt, 1984; Wickens, 1986). Indeed, there is substantial evidence to suggest that object displays are superior to alphanumeric displays in many applications where identification of system state requires integrating data from a number of information sources (Casey, 1986; Carswell and Wickens, 1987; Wickens, 1986).

The superiority of object displays has been attributed, in part, to the perceptual cues and redundant information inherent in such representations (Garner, 1974, 1980; Pomerantz, Pristach and Carson, 1987). The redundancy and perceptual cues can be used by an operator to simplify classification of system data by associating a unique object configuration or set of salient features to a specific system-state category. The mapping of objects or features to a state category occurs when the values of system variables are correlated with a particular state category, and the physical representation of those values of system variables creates a configuration with a unique size, shape or orientation. Consequently, the operator need not attend to specific values of system variables, but can rely on rapid, possibly holistic, integral processing of an object-like configuration or set of salient features to determine the state of the system. In terms of multiple-resource theory (Wickens, 1984), object displays produce a spatial code that allows integral processing of system data.

Alphanumeric displays and tabular formats, on the other hand, require the operator to attend to each individual system variable and serially process system data as a verbal code. Since separable displays require mental manipulation of numerical values to determine category membership, the underlying correlational structure of system data (as defined by Garner, 1974) is not as readily apparent and, presumably, requires more processing time than a more integral display. Consequently, many researchers have concluded that the appropriate display format of system data is dependent upon the underlying statistical properties of data in a task (Goldsmith and Schvaneveldt, 1984; Wickens, 1986).

In general, however, previous research has been equivocal on the exact nature and form of the relationship between the statistical properties of data in a task and its physical representation. Goldsmith and Schvaneveldt

(1984) have shown that an integral representation (an object display) is consistently superior to a separable representation (a bargraph display) in a multi-cue criterion-judgment task. Casey (1986), on the other hand, has shown that a bargraph display is superior to an object (polygon) display when the operator is required to focus attention on a single, critical dimension in a detection and diagnosis task. Carswell and Wickens (1987) have further demonstrated that an object display (in the form of a triangle) is superior to a separable display (a bargraph) when integration of data is required, but that the superiority disappears when the task requires focusing attention on a single data source. None of those studies, however, directly addressed identification of system state as a categorization process.

In addition, indirect evidence suggests that tasks that require focused attention on a single dimension or attribute of displayed information are best represented by a separable display (Garner, 1974; Wickens, 1986), especially when the task is primarily composed of detection or straightforward decision making (Triesman and Gelade, 1980; Kahneman and Triesman, 1984). In more complex situations, however, the need for both types of display formats may be evidenced (Coury and Pietras, 1989).

The Impact of Uncertainty

Unfortunately, none of the previously cited studies has directly addressed the fact that identification of system state is a categorization process; none has specifically considered the importance of uncertainty in the selection of displays. The first issue is primarily concerned with the nature and characteristics of the operator's task and is formulated in terms of Rasmussen's model of the operator.

The second issue relates directly to object display superiority and integral processing of system data. The underlying uncertainty in the statistical properties of a system-state identification may determine when the advantages of an object display for presenting multidimensional, correlated data may be nullified. For instance, if two fundamentally different system states exhibit very similar diagnostic cues, the object displays for those two states may be very similar in appearance. In such a situation, the classification task becomes primarily one of discrimination, requiring the operator to focus on specific values of system variables to distinguish between the two states. Consequently, the superiority of an object display may diminish when the state of the system is uncertain and identification of system state requires that the display be decomposed into its individual system variable values.

Few studies have directly addressed the issue of uncertainty, or systematically manipulated uncertainty to determine the conditions under which separable and integral representations of correlated data are superior. In fact, it appears quite possible that much of the research that has evaluated the merits of integral and separable displays has failed to control for the effects of uncertainty (as defined in this research).

Research Purpose and Rationale

The purpose of the experiments reported in this paper was to explore the effects of uncertainty, and evaluate the impact of uncertainty on classification performance when correlated, multidimensional system data are presented in either integral or separable form. Three types of displays were used in these experiments: a Configural display; a Digital display; and a Bargraph display.

The Configural display (shown in Figure 1a) presents instances of system data in integral form. The Configural display maps the values of system variables defining a system state onto a single object-like configuration (a polygon). The display allows the operator to classify system state by attending to the overall shape or configuration of the display without attending to any specific value of a process variable. This allows the operator to process the data as a spatial code.

The Digital display (shown in Figure 1b) presents the same system variables as independent values (i.e., separate digits). The Digital display is a separable representation that maximizes the separation between process variables, places emphasis on verbal coding of the data, and minimizes the opportunity for configural properties to emerge.

The Bargraph display (shown in Figure 1c) is used in this research as a display format that can possess both separable and configural properties. Numerous researchers have advocated the use of bargraphs as a separable display because each bar independently presents each variable of interest (Wickens, 1986; Casey, 1986; Carswell and Wickens, 1987). In addition, the Bargraph display is amenable to verbal coding; e.g., the display shown in Figure 1c can be coded as "high-low-low-high". In fact, a recent industry guide for designing computer-generated displays recommends bar charts for unidimensional data and comparison tasks, and discourages their use for multidimensional data in status and pattern recognition tasks (Frey and Sides, 1984, pg. 5-3). Such interpretations ignore, however, the real possibility that the Bargraph display can possess object-like properties. When used in correlated tasks, the heights of the bars produce contours unique to a system-state category, and the operator need only attend to the overall shape of the Bargraph configuration to classify system state. Two recent studies have reported on the configural properties of the bargraph display (Coury and Purcell, 1988; Buttigieg, Sanderson and Flach, 1988).

INSERT FIGURE 1a, 1b AND 1c

ABOUT HERE

To assess identification of system-state as a categorization process, a classification task using four system-state categories was constructed. Our objective was twofold: create a task that was consistent with our model of operator decision making; emphasize the essence and cognitive requirements of the decision problems faced by an operator, rather than directly simulate an actual system.

Consequently, a task was constructed that met two basic requirements: 1) there be a set of attributes and values of those attributes that would map an instance of system data onto a specific state category; and 2) that specific instances vary in uncertainty (i.e., diagnosticity). In addition, the demands of the task would balance fidelity and learning; the decision rules for accurate classification would not be trivial, but would, at the same time, allow the task to be learned by untrained people in a reasonable amount of time.

The task used in these experiments defined system-state categories as ranges of values along four dimensions. Each dimension represented a system variable and a specific range of values for each system variable was combined to define a system-state category. The range of values defining each of the four system states is presented in Table 1 (Q, M, B and H are the labels for the four system variables); overlapping values provided nondiagnostic values for the process variables. That is, uncertainty was introduced by creating an overlap (i.e., borderline condition) between state categories in which one or more system variables could take on values that simultaneously define more than one system state.

The borderline region represents an area of uncertainty about the identity of an instance of system state. One might expect, then, that as any given instance of system data approached the borderline condition, the operator would need more precise information about that instance (i.e., have to attend more closely to specific values of system variables). In such a situation, a display that can be easily partitioned into its individual components (the Digital or Bargraph displays) may be superior to a display that masks subtle differences between values of state variables (the Configural display).

Conversely, when system-state is certain (i.e., when an instance of system data is most characteristic of a specific state category), then a display that allows integral processing of system data should be superior. In these experiments, one can predict that the perceptual cues inherent in the Configural display provide the most apparent association of system variables to state categories in situations of low uncertainty and high diagnosticity. One might also expect that the Bargraph display, because of its integral processing potential, to produce results equivalent to the Configural display, or at the very least, intermediate results depending on the extent to which operators use a integral or separable processing strategy for that display.

Two experiments were conducted to evaluate the impact of integral and separable displays of system data on classification performance. The primary purpose of Experiment 1 was to determine the extent to which uncertainty and diagnosticity affects the processing of integral and separable displays of system data. Experiment 1 was also concerned with the effect of extended practice on classification performance, since there is indirect evidence suggesting that extended practice may result in very efficient processing of information in classification tasks (Garner, 1974).

In the training session, operators were required to attain a prespecified classification accuracy criterion. The criterion was set to ensure that the task was well learned and performance during the test sessions would not be contaminated by inadequate understanding of the decision rules defining category membership. Only those operators who reached criterion were allowed to participate in all experimental sessions. This appeared to be a reasonable performance expectation since actual operators are typically highly trained and very familiar with the system.

The purpose of Experiment 2 was to determine if additional exposure during learning to instances of high uncertainty would improve classification performance. The uniform distribution of instances of system state in Experiment 1 ensured that operators had equal exposure to the the full range of category membership. It is possible, however, that an operator's ability to effectively discriminate between state categories when uncertainty is high requires that the operator have had considerable exposure to instances of system data that are low in diagnosticity. Without sufficient experience, one might expect performance on instances of high uncertainty to be relatively worse than performance with the very certain instances, with the differences between display types to be an artifact of learning. Consequently, the distribution of instances of system-state in Experiment 2 was biased toward the area of uncertainty within each state category; i.e., operators were exposed to more instances of high uncertainty during learning and extended practice in Experiment 2 than in Experiment 1.

METHODS: Experiment 1

Subjects

Thirty subjects, ranging in age from 18 to 40, participated in both the training and extended practice sessions of Experiment 1. They were drawn from the University of Massachusetts undergraduate and graduate student population and the local community, and they were paid $5.00 per hour for participation. These subjects represent only those who reached a criterion level of classification performance during training and were allowed to complete the entire experiment.

System State Categories

System state categories were defined in this experiment (as well as in Experiment 2) as specific ranges of values across system variables. These ranges of values, along with the four system-state categories are shown in Table 1.

The experimental stimulus set was comprised of 256 separate instances of system data, with an equal number of instances from each of the four state categories. The correlational structure of the state categories is similar to a binary decision tree. State Categories 1 and 3 are characterized by the same ranges of values along variables B and H, as are State Categories 2 and 4. Thus, information provided by values for variables B and H systematically reduces the the possible system states to two. The additional information provided by the values of variables Q and M can be used to determine the actual state category of an instance.

As explained in the Introduction, uncertainty was created with a borderline, overlapping region where instances of system state could belong simultaneously to two state categories. The borderline condition represents the transition of one system state to another. For example, an instance of system data with values of 50, 50, 80 and 20, for variables Q, M, B, and H, respectively, is possible in both System States 2 and 4 (but not 1 or 3), and a response of State 2 or State 4 would be correct. Thus, the borderline condition creates an area of overlap between two state categories where the values of the two most critical variables are no longer diagnostic. The six steps from the borderline were created by systematically manipulating the diagnosticity and uniqueness of the values of process variables for a state category. At Step 1, at least one of the critical variables was nondiagnostic and the other variable was diagnostic (but set at a value close to the range of values found in the overlapping system state). Steps 2 through 6 manipulated the similarity of the process variables to the overlapping system state. As distance from the borderline increased from Step 2 to Step 6, the values of the process variables became increasingly different from both the overlapping and non-overlapping state categories, producing combinations of process variables that were unique to that state category.

System State Representations

Three types of display formats were used to present system data in this experiment: an integral representation (the Configural display); a separable representation (the Digital display); and a representation with the potential for both separable and integral properties (the Bargraph display). Examples of the representations are presented in Figures 1a, 1b and 1c.

Experimental Tasks

The subjects acted as "operators" whose task was to identify system state. In this experiment, identification of system state required classifying the displayed instances of system data into one of four state categories. The classification scheme required operators to integrate information from the four system variables presented to them, and learn the decision rules defining state categories.

Procedure

The experiment was divided into two sessions: training, and extended practice. The classification learning technique employed by Coury and Drury (1986) was adapted for use in this experiment.

Subjects were randomly assigned to one of the three display groups. Without describing the underlying classification scheme, the experimenter reviewed with each subject the procedures, demonstrated the task, and explained the information provided by the feedback monitor. The feedback monitor provided information regarding the accuracy of the response, the correct system state, and variable-specific information about state category membership. At no time, however, did the subject receive specific information about the possible range of values for each state category.

Training. During training, the "operators" classified the 256 instances of system data. Presentation of each instance constituted a training trial. Each training trial followed the same pattern: presentation of an instance of system data; the operator's response; and feedback on the accuracy of the response. The operator indicated his or her response by pressing one of four keys on the keyboard corresponding to a state category. The cycle was repeated until each subject had viewed all 256 trials in the stimulus set.

Extended Practice. After a short break following the training session, operators who had reached criterion returned to perform the second classification session. In this session, instances were presented to operators using the same display format as in the training session. The stimulus set was the same as used in the training session, but presented in a different random order.

The classification task in both the Training and Extended Practice sessions was self-paced. Each trial was initiated by the operator, with the instance of system data remaining displayed until the operator responded. There was no restriction on the amount of time the feedback monitor could be viewed during the Training session. A DEC Pro 300 series microcomputer was used to display stimuli and collect operator response data. The stimuli were presented on a high resolution graphics, monochrome display, with a second monitor used to present feedback to operators.

Data Measurement

For each operator, response time and accuracy data for each trial in both sessions was recorded. Reaction times were measured as the interval between the onset of an instance of system data and the operator's response. Only correct responses were used for analysis. Accuracy was defined as the proportion of instances of system data correctly classified. Operators were trained to a prespecified criterion. The criterion was 90% correct responses in the last 100 trials of the training session.

Both reaction time and percent correct were summarized by averaging across responses on 32 trials; thus, there were eight consecutive blocks of data for each session. To evaluate the effect of uncertainty, an analysis was performed on the response times for instances of system data occupying predetermined incremental distances from the borderline condition. The instances selected for analyses were at the borderline and the six steps away from the borderline condition (as described in the System State Category section); the six steps represented equal step sizes across the range of category membership. Only times for correct responses were analyzed and the results were averaged across state categories.

Experimental Design and Analysis

The experiment was a multifactor repeated measures design with operators nested under display type. Display type was the between-subjects variable and Blocks of Trials were within-subject variables. Each of the dependent measures for each session were subjected to a multifactor ANOVA. All factors except operators were treated as fixed. A second ANOVA evaluated the effect of uncertainty as a within subjects variable.

3. RESULTS: Experiment 1

Accuracy and Trials to Criterion

The operator's ability to accurately classify instances of data rapidly improved during the first half of the Training session, reaching an asymptotic level by the end of the fourth block of trials. By the end of Training and throughout the Extended Practice session accuracy stabilized at a level greater than the prescribed criterion. The proportion of instances correctly classified for each block of 32 trials for the three types of displays in the Training and Extended Practice sessions is presented in Figure 2.

INSERT FIGURES 2a AND 2b

ABOUT HERE

The ANOVA of these data found no significant difference between display types in the Training session, $F(2,26) = .15$, $p = .862$. There was, however, a significant difference between display types in the Extended Practice session, $F(2, 26) = 4.66$, $p = .0186$, although the difference in mean accuracy between the best display (the Bargraph) and the worst display (the Configural) was only 3.8 percent and all displays were well above criterion. The significant improvement in classification accuracy in the Training session was reflected in the main effect of Blocks of Trials, $F(7,14) = 48.71$, $p < .0001$. In the Extended Practice session, accuracy had stabilized above criterion and the main effect of Blocks of Trials was no longer significant, $F(7,14) = .89$, $p = .517$. All other main effects and interactions were not significant.

Response Times: Training and Extended Practice

The Bargraph display produced the fastest response times in both the Training and Extended Practice sessions; response times for the Digital display group were slowest, with the Configural display producing intermediate results. In addition, there was a major reduction in response times across display types during the Training session, with a less pronounced reduction in response times occurring in the Extended Practice session. Mean response times for the three display types for both the Training and Extended Practice sessions are presented in Figure 3.

INSERT FIGURES 3a AND 3b

ABOUT HERE

The ANOVA of the Training session response time data showed that the effect of Display type to be significant, $F(2,26) = 4.68$, $p = .0184$, and Blocks of Trials to be highly significant, $F(7,14) = 37.02$, $p < .0001$. The ANOVA of the Extended Practice session showed the effect of Display type to be highly significant, $F(2,26) = 6.67$, $p = .005$, and Blocks of Trials to be highly significant, $F(7,14) = 6.83$, $p < .0001$.

Distance from Borderline

The previous analyses did not consider the impact of uncertainty on classification performance. To evaluate the effect of uncertainty, an analysis was performed on accuracy and response times to instances drawn from the borderline and the six step sizes described in the Methods sections.

Classification accuracy for the three types of displays was found to be affected by uncertainty (as defined by Distance from the Borderline). Operators made the greatest number of classification errors under conditions of high uncertainty with greatest accuracy occurring under conditions of low uncertainty. There was, however, no significant difference in classification accuracy among the three display types. Mean accuracy for each display type as a function of Distance from Borderline is presented in Table 2. The ANOVA of these data found only the main effect of Distance from Borderline to be significant, $F(6,102) = 8.38$, $p < .001$; all other main effects and interactions were not significant.

Mean response times for the three types of displays as a function of Distance from the Borderline are presented in Figure 4. Uncertainty (as defined by Distance from the Borderline) had a significant effect on all three display types; response times were slowest under conditions of high uncertainty, and decreased monotonically as certainty increased. The ANOVA of these data found the main effect of Display type to be significant, $F(2,26) = 7.52$, $p = .0027$, as well as the Display type by Distance from Borderline interaction, $F(12,156) = 1.93$, $p = .0341$. Analysis of the simple main effects of Distance from Borderline for each of the display types found the effect of uncertainty to be highly significant in the following way: Bargraph display, $F(6,54) = 14.07$, $p < .0001$; Configural display, $F(6,48) = 11.16$, $p < .0001$; and the Digital display, $F(6,54) = 4.49$, $p < .001$.

INSERT FIGURES 4 and 5

ABOUT HERE

Contrasts between display types were found to be significant. Analysis of the difference between the Configural display and the Bargraph display revealed a significant simple main effect of Display type, $F(1,17) = 5.79$, $p = .0278$, but no Display type by Distance from Borderline interaction, implying that the effect of uncertainty was additive for the Configural and Bargraph displays. The same analysis for the Configural and Digital display found the Display type by Distance from the Borderline interaction to be significant, $F(6,102) = 2.59$, $p = .0223$.

The interaction between the Configural and the Digital display appears to be due to the differential effect of uncertainty on the two display types; the Configural display produced significantly faster response times than the Digital display under conditions of low uncertainty. As uncertainty increased, however, the superiority of the Configural display diminished, with the Digital display becoming superior to the Configural display near the borderline condition.

DISCUSSION: Experiment 1

The results of this experiment clearly demonstrate the superiority of the Bargraph display in a multidimensional, classification task. The Bargraph was, however, susceptible to the effects of uncertainty in much the same way as the Configural and Digital displays. In all three cases, response times significantly increased as uncertainty increased. In addition, these results clearly showed the conditions under which the Configural display would be superior to the Digital display. When an instance was unmistakably a member of a particular category, an object-like representation (the Configural display) clearly enhanced the operator's ability to process system data and classify system state. Once uncertainty reached a certain point and the diagnosticity of critical variables diminished, a more precise representation of system data was necessary and the Digital display became superior.

METHODS: Experiment 2

The methods and procedures used in Experiment 2 were the same as in Experiment 1 except for a change in the distribution of instances across category membership.

Subjects

Thirty subjects participated in both the Training and Extended Practice sessions of Experiment 2. They were drawn from the University of Massachusetts undergraduate and graduate student population and the local community, and they were paid $5.00 per hour for participation. These subjects represent only those who reached criterion in the Training session and were allowed to complete the entire experiment. The subjects ranged in age from 18 to 40.

System State Categories and Representations

System state categories used in this experiment were the same as those used in Experiment 1 (see Table 1). In Experiment 2, subjects were exposed to a greater number of instances from the borderline and near borderline areas of a state category; this was accomplished by removing instances from areas of low uncertainty in each of the state categories. The purpose of this manipulation was to provide the subject with more experience classifying instances of high uncertainty/low diagnosticity. Consequently, the stimulus set in this experiment was comprised of 224 separate instances of system data, with an equal number of instances from each of the four state categories. The instances were presented to subjects using the displays used in Experiment 1.

Task and Procedure

The same task and procedure used in Experiment 1 was used in Experiment 2. Each subject participated in a Training and Extended Practice session, with each session comprising 224 instances of system state. As in Experiment 1, summarized data were submitted to ANOVA. In the Distance from Borderline analysis, the biased distribution of instances of system state resulted in response times occurring only in the borderline condition and in step sizes 1, 2, 4 and 6.

RESULTS: Experiment 2

Classification Accuracy

The operator's classification accuracy increased rapidly during the first half of the training session, reaching asymptotic levels for all display types by the end of the fifth block of trials (Figure 6). Classification accuracy data for the Training and Extended Practice Sessions in Experiment 2 are presented in Figure 6. The ANOVA showed that the main effect of Blocks of Trials was highly significant during the Training session, $F(6,156) = 91.42$, $p < .0001$. This pattern during the Training Session was similar to the pattern found in Experiment 1.

INSERT FIGURE 6a AND 6b

ABOUT HERE

Classification accuracy was greatest for the Digital display with the Configural and Bargraph display producing comparable performance, resulting in a main effect of Display type approaching significance, $F(2,26) = 2.98$, $p = .068$. By the last block of trials of training, however, the differences between the most accurate and the least accurate display was only 1.2 percent. Although classification accuracy with the Digital display appears to reach criterion before the other displays (by the third block of trials rather than the fifth block), the Display Type by Blocks of Trials interaction is not significant, $F(12,156) = .70$.

By the end of Training and throughout the Extended Practice session accuracy remained constant at a level greater than the prescribed criterion.

The ANOVA of the accuracy data from the Extended Practice session found no significant main effect of Display type or Blocks of Trials, and no Display Type by Blocks of Trials interaction.

Response Times: Training and Practice

The greatest reduction in response times across trials was found during the Training session (see Figure 7). Response times did, however, continue to decline in the Extended Practice session, albeit at a much slower rate. The ANOVAs of these data revealed a highly significant main effect of Blocks of Trials in the Training session, $F(6,156) = 33.89$, $p < .0001$, and in the Extended Practice session, $F(6,156) = 7.94$, $p < .0001$. The trend was the same for all three displays since no significant effect of display type or significant Display type by Blocks of Trials interaction was found in either the Training session or in the Extended Practice session.

INSERT FIGURES 7a AND 7b

ABOUT HERE

Distance from Borderline

To evaluate the effect of uncertainty in Experiment 2, an analysis was performed on classification accuracy and response times to instances occupying the borderline and four step sizes from the borderline described in the Methods section. Analysis of the accuracy data (presented in Table 2) found no significant differences between display types or interactions with Distance from the Borderline.

Response times were, however, affected by uncertainty. In general, the response times to the Configural display were slowest of the three displays, with the Bargraph and Digital displays producing very similar response times. Mean response times for the three display types as a function of Distance from the Borderline are presented in Figure 5.

All three display types were significantly affected by uncertainty; operators produced their fastest response times under conditions of low uncertainty, and their slowest response times under conditions of high uncertainty. Analyses of the simple main effects of Distance from the Borderline were significant for all three displays: Bargraph, $F(4,36) = 6.62$, $p < .0005$; Configural, $F(4,36) = 7.10$, $p < .0005$; and Digital, $F(4,32) = 7.46$, $p < .0005$.

The displays appeared not to be affected by uncertainty in the same way; the Display type by Distance from Borderline interaction was found to be significant, $F(8,104) = 2.32$, $p < .05$. The response times for the Digital and Bargraph displays appear to be quite similar; indeed, the simple comparison of the Digital and Bargraph display found no significant difference between the two displays, $F(1,17)) = .02$, or an interaction with Distance from Borderline, $F(4, 68) = 1.17$. Thus, the source of the interaction appears to be primarily due to the Configural display being affected differently by uncertainty relative to the Digital and Bargraph displays.

The relative change in response times from Experiment 1 to Experiment 2 indicates the effect of biasing the distribution of instances on uncertainty. The Digital display was least affected; response times as a function of uncertainty in this experiment were within 311 msecs of the equivalent response times obtained in Experiment 1. This was not the case for the Bargraph display nor the Configural display; the Bargraph display exhibited the greatest increase in response times (1160 msecs), while the Configural display was less affected (978 msecs).

DISCUSSION: Experiment 2

Comparison of the results from Experiment 1 and 2 shows the effect of a biased distribution of instances. Exposure to more uncertain instances during learning appears to have had little effect on the Digital group's ability to process instances of system data; i.e., the response times to the Digital display in Experiment 1 were virtually identical to the response times produced in Experiment 2. These results suggest that a separable representation of correlated data is less affected by the distribution of instances of system-state.

The biased distribution had the greatest impact on the Bargraph and Configural displays. The mean response times at all levels of uncertainty are considerably slower in Experiment 2 than in Experiment 1. Thus, greater exposure to instances possessing some degree of uncertainty appears to either diminish the operator's ability to use the configural properties in the two displays, or changes the operator's processing strategy. Although these results may be simply due to the differences between subjects in the two experiments, the difference is great enough to warrant discussion and to motivate replication in future research.

GENERAL CONCLUSIONS

The results of this research demonstrate two very important effects of uncertainty on the identification of system state and the selection of displays for multidimensional data. First, there was a positive relation between uncertainty and the time required to classify an instance of system data. Second, the superiority of a particular display format varied as a function of uncertainty. In general, displays possessing configural and object-like properties were processed faster when uncertainty was low, while displays possessing separable properties were processed faster when uncertainty was high. Specific deviations from those general findings are discussed below.

Uncertainty, as defined in this research, is related to the mapping of system data to system-state categories, and has a significant impact on operator performance. In general, as uncertainty increases, the time required to classify an instance of system data increases. When uncertainty is low and an instance of system data is uniquely characteristic of a particular state category, the categorization process can occur relatively quickly. Conversely, when uncertainty is high and values of process variables cannot be mapped to a single system-state category, the categorization process requires more time.

Uncertainty, then, emerges as an important factor in display design, and can be used to determine the relative merits of display formats. When an instance of system data is unmistakably a member of a particular system state, displays that possess object-like properties or emergent features (the Configural and the Bargraph display) can enhance the operator's ability to quickly process system data and accurately identify the state of a system. The advantage of such displays diminishes, however, as uncertainty increases or experience with the full range of category membership is not sufficient (as in Experiment 2).

Once uncertainty reaches a certain point and the diagnosticity of critical variables diminishes, a more precise presentation of system data (a separable display) is necessary. Under conditions of high uncertainty, the Configural display appears to be difficult to decompose, thereby reducing the operator's ability to detect subtle, but critical, changes in individual system variables, resulting in an increase in response times. Only the precise values presented by the Digital display and the separable qualities of the Bargraph display allowed the subtle variations to become quickly apparent to the operator in these experiments. This is consistent with previous display research (e.g., Carswell and Wickens, 1987) and studies concerned with focused and divided attention (Kahneman and Triesman, 1984; Triesman and Gelade, 1980). From a practical standpoint, then, a separable display format appears to enhance the identification of system state when uncertainty is high.

It would be tempting to conclude that the relative differences in superiority among the three types of displays were due to changes in the salience of certain perceptual cues in the display under conditions of uncertainty. Such a conclusion would be consistent with Pomerantz's emergent features approach (Pomerantz, Pristach and Carson, 1987), but would ignore other competing explanations based on a configural approach (Garner, 1980) or the principle of display proximity proposed by Carswell and Wickens (1987).

Unfortunately, definitive support for a singular approach to assessing display formats in a multidimensional classification task is not readily apparent. For instance, Garner (1974) describes configural properties in much the same terms as Pomerantz's emergent features (Pomerantz, Pristach and Carson, 1987). Sanderson, Flach, Buttigieg and Casey (1989) have also shown that an emergent features approach can predict performance in a failure detection task as well as the principle of display proximity proposed by Carswell and Wickens. In addition, the way people process displays may be mediated by individual differences; Purcell and Coury (1988) have argued that operators do not necessarily respond to configural and separable displays in the same way and may, in fact, adopt a number of different processing strategies.

One can conclude from these experiments that salient perceptual cues in a display are a function of the underlying statistical properties of a task. When the contingencies among variables in a task combine in such a way as to define a specific response (i.e., a state category), there arises an opportunity for display elements to combine and emerge as useful classification cues. Even the operators using the Digital display were able to take advantage of certain patterns in the values of process variables to reduce their response times when uncertainty was low. Thus, uncertainty provides a means for operationally defining the underlying statistical properties of a task and revealing the conditions under which elements of a display can enhance operator performance.

The results from Experiment 2 further demonstrate the effects of uncertainty on operator performance. In that experiment, operators were faced with a preponderance of uncertain instances. If the effects of uncertainty revealed in Experiment 1 were operating in Experiment 2, then one would expect that a display allowing the use of a separable processing strategy would be most effective for classifying instances of system state. Furthermore, if one accepts the assumption that the Digital display is processed as a serial, verbal code, and the assumption that the Bargraph display can be decomposed and coded in a serial fashion, then one might expect those displays to enhance classification performance under conditions of high uncertainty. The Digital and Bargraph displays in Experiment 2 produced the best (as well as equivalent) performance under conditions of uncertainty, implying that those displays were being processed in a separable fashion (one cannot conclude, however, that the Bargraph was being processed as a verbal code). It is also interesting to note that the Configural display was the most adversely affected by the distribution of uncertainty, indicating that the superiority of the Configural display is evident only under conditions of low uncertainty.

Notice, too, that all three display formats produced equivalent classification accuracy. Although the Configural display produced slightly better accuracy in Experiment 1, it is unlikely that such a difference would be important in an operational setting. It is also interesting to note that all of the operators reached the same level of accuracy by the end of training sessions, and there was little variation in the rate of learning among displays. The results indicate that sufficient training can effectively eliminate certain differences among displays in the early stages of learning.

These experiments also demonstrate that the configural and separable properties of the Bargraph display is related to uncertainty. The superiority of the Bargraph display in Experiment 1 indicates that this type of display can possess the properties of an object display, and can be used to enhance classification performance, especially when uncertainty is low. When uncertainty is high and a separable processing strategy is necessary for effective performance (as in Experiment 2), then the Bargraph display can be processed separably and result in better performance than the Configural display and equivalent performance to the Digital display. One can conclude, then, that the results of this research add to the evidence presented in Buttigieg, Sanderson and Flach (1988) and Coury and Purcell (1988) that the Bargraph display can possess both integral and separable properties. From a display design point of view, the results also suggest that the Bargraph display may be a very versatile format and is potentially the best choice for displaying multidimensional, correlated data when the range of uncertainty is large.

In conclusion, it is important to reiterate the importance of uncertainty, and the impact of such a factor on operator performance and the superiority of a particular display format. In terms of display design, a principle seems to emerge: when instances of system data are unmistakably a member of a state category, a display allowing configural properties or emergent features to emerge will enhance the identification of system state. When uncertainty is high, however, the precise presentation of system data provided by a separable display is necessary. It is interesting to note that without the analysis of uncertainty, the differences between displays in these experiments would be, for all practical purposes, not significant. Consequently, these results suggest that uncertainty can have a profound impact on the processing of multidimensional, correlated system data.

ACKNOWLEDGEMENTS

The authors would like to take this opportunity to thank our contract monitor, V. Grayson Cuqlock-Knopp, at the U.S. Army Human Engineering Laboratory, Aberdeen Proving Grounds, MD, for her support, encouragement and input during the pursuit of this research. Without her continued support, this research would not be possible. We are also indebted to the Department of Industrial Engineering and Operations Research at the University of Massachusetts for providing the facilities necessary for conducting this research.

REFERENCES

Barnett, B.J., and Wickens, C.D. (1988), Display proximity in multicue information integration: the benefit of boxes. Human Factors, 30 (1), 15-24.

Boulette, M.D., Coury, B.G., and Bezar, N.A. (1987), Classification of multidimensional data under time constraints: evaluating digital and configural display representations. In the Proceedings of the Human Factors Society 31st Annual Meeting (Santa Monica, CA: Human Factors Society), 116-120.

Buttigieg, M., Sanderson, P.M., and Flach, J. (1988), Object vs separate displays for process failure detection: the emergent features approach. In the Proceedings of the Human Factors Society 31st Annual Meeting (Santa Monica, CA: Human Factors Society), 1340-1344.

Carswell, C.M., and Wickens, C.D. (1987), Information integration and the object display: an interaction of task demands and display superiority. Ergonomics, 30(3), 511-527.

Casey, E.J. (1986), Visual display representation of multidimensional systems: the effect of information correlation and display integrality. In the Proceedings of the Human Factors Society 30th Annual Meeting (Santa Monica, CA: Human Factors Society), 430-434.

Coury, B.G., and Drury, C.G. (1986), The effects of pacing on complex decision-making inspection performance. Ergonomics, 29(4), 489-508.

Coury, B.G. and Pietras, C.M. (1989), Digital and graphic displays for process plant control. Ergonomics, in press.

Coury, B.G. and Purcell, J.A. (1988), The bargraph as a configural and separable display. In the Proceedings of the Human Factors Society 32nd Annual Meeting, Vol. 2 (Santa Monica, CA: The Human Factors Society), 1361 - 1365.

Coury, B.G., Boulette, M.D., Zubritzky, M.C., and Fisher, D.L. (1986), Classification of multidimensional data: evaluating displays to aid identification of system state. In the Proceedings of the 1986 IEEE International Conference on Systems, Man, and Cybernetics, Vol. 2 (Piscataway, NJ: IEEE Service Center), 1503-1508.

de Kleer, J. and Brown, J.S. (1983), Assumptions and ambiguities in mechanistic mental models. In D. Gentner and A.L. Stevens (Eds.), Mental Models. Hillsdale, NJ: Lawrence Erlbaum.

Frey, P.R., and Sides, W.H. Jr. (1984). Computer-Generated Display System Guidelines. Vol. 1: Display Design. EPRI Report No. NP-3701. Palo Alto, CA: Electric Power Research Institute.

Garner, W.R. (1974), The Processing of Information and Structure. New York: John Wiley & Sons.

Garner, W.R. (1980), Functional aspects of information processing. In R. Nickerson (Ed.), Attention and Performance VIII. Hillsdale, NJ: Lawrence Erlbaum.

Goldsmith, T.E. and Schvaneveldt, R.W. (1984), Facilitating multiple-cue judgments with integral information displays. In J.C. Thomas and M.L. Schneider (Eds.), Human Factors in Computer Systems. Norwood, NJ: Ablex Publishing Corp.

Green, D.M., and Swets, J.A. (1966), Signal Detection Theory and Psychophysics. Huntington, NY: Robert E. Krieger.

Henneman, R.L. (1988), Human problem solving in dynamic environments: understanding and supporting operators in large-scale complex systems. In W.B. Rouse (Ed.), Advances in Man-Machine Systems Research, Vol. 4. Greenwich, CT: JAI Press Inc.

Hendler, J.A. (1988), Expert Systems: The User Interface. Norwood, NJ: Ablex Publishing Corp.

Holland, J.H, Holyoak, K.J., Nisbett, R.E., and Thagard, P.R. (1986), Induction: Processes of Inference, Learning, and Discovery. Cambridge, MA: The MIT Press.

Jacob, R.J.K, Egeth, H.E., and Bevan, W. (1976), The face as a data display. Human Factors, 18, 189-200.

Kahneman, D. and Triesman, A.M. (1984), Changing views of attention and automaticity. In R. Parasuraman and D.R. Davies (Eds.), Varieties of Attention. New York: Academic Press, Inc.

Miller, R.A. (1985), A systems approach to modeling discrete control performance. In W.B. Rouse (Ed.), Advances in Man-Machine Systems Research Vol. 2. Greenwich, CT: JAI Press.

Nickerson, R.S. (1986), Using Computers: Human Factors in Information Systems. Cambridge, MA: The MIT Press.

Norman, D.A., and Draper, S.W. (1986), User Centered System Design. Hillsdale, NJ: Lawrence Erlbaum Assoc.

Olson, J. Reitman (1987), Cognitive analysis of people's use of software. In J. Carroll(Ed.), Interfacing Thought: Cognitive Aspects of Human-Computer Interaction. Cambridge, MA: the MIT Press.

Pomerantz, J.R., Pristach, E.A., and Carson, C.E. (1987), Attention and object perception. In B. Shepp and S. Ballesteros (Eds.), Object Perception: Structure and Processes. Hillsdale, NJ: Lawrence Erlbaum.

Purcell, J.A., and Coury, B.G. (1988), Asymmetric transfer of training between integral and separable displays. In the Proceedings of the Human Factors Society 32nd Annual Meeting, Vol. 2 (Santa Monica, CA: The Human Factors Society), 1366 - 1370.

Rasmussen, J. (1986), Information Processing and Human-Machine Interaction. New York: North-Holland.

Rasmussen, J. (1988), A cognitive engineering approach to the modeling of decision making and its organization in process control, emergency management, CAD/CAM, office systems, and library systems. In W.B. Rouse (Ed.), Advances in Man-Machine Systems Research Vol. 4. Greenwich, CT: JAI Press, Inc.

Sanderson, P.M., Flach, J.M., Buttigieg, and Casey, E.J. (1989), Object displays do not always support integrated task performance. Human Factors, 31(2), 183-198.

Shneiderman, B. (1987), Designing the User Interface: Strategies for Effective Human-Computer Interaction. Reading, MA: Addison-Wesley.

Smith, E.E., and Medin, D.L. (1981), Categories and Concepts. Cambridge, MA: Harvard University Press.

Sprague, R.H., and Carlson, E.D. (1982), Building Effective Decision Support Systems. Englewood Cliffs, NJ: Prentice-Hall.

Triesman, A.M. and Gelade, G. (1980), A feature integration theory of attention. Cognitive Psychology, 12, 97-136.

Wickens, C.D. (1984), Engineering Psychology and Human Performance. Columbus, OH: Charles E. Merrill.

Wickens, C.D. (1986), The Object Display: Principles and a Review of Experimental Findings. Technical Report No. CPL 86-6. Champaign, ILL: Cognitive Psychophysiology Laboratory.

Woods, D.D., Wise, J.A., and Hanes, L.F. (1981), An evaluation of nuclear power plant safety parameter display systems. In R.C. Sugerman (Ed.), Proceedings of the Human Factors Society 25th Annual Meeting. Santa Monica, CA: The Human Factors Society.

Table 1
Ranges of values for System Variables

System State	Q	M	B	H
	System Variable			
1	25-51	49-75	0-26	74-100
2	25-51	49-75	74-100	0-26
3	49-75	25-51	0-26	74-100
4	49-75	25-51	74-100	0-26

Table 2
The Effects of Uncertainty on Classification Accuracy

Experiment 1

		Distance from Borderline Step Size					
	Border	1	2	3	4	5	6
Digital:	.96	.89	.97	.96	.96	.94	.96
Bargraph:	.97	.96	.95	.96	.97	.96	.96
Configural:	.99	.84	.97	.94	.97	.96	.95

Experiment 2

		Distance from Borderline Step Size			
	Border	1	2	4	6
Digital:	.96	.97	.94	.95	.95
Bargraph:	.95	.97	.98	.93	.89
Configural:	.99	.98	.95	.91	.93

Figure Captions

Figures 1a, 1b and 1c. The three types of displays used in Experiments 1 and 2. Figure 1a shows an instance of system data presented as the Configural display; Figure 1b shows an instance of system data presented as the Digital display; and Figure 1c shows an instance of system data presented as the Bargraph display. In all cases the values of the system variables are the same.

Figure 2. Mean proportion correct in Experiment 1 for the Bargraph, Digital and Configural displays as a function of blocks of trials during Training and during the Extended Practice sessions.

Figure 3. Mean response times (in msecs) in Experiment 1 for the Bargraph, Digital and Configural displays as a function of blocks of trials during Training and during the Extended Practice sessions.

Figure 4. Mean response times from Experiment 1 for the Bargraph, Digital and Configural display groups as a function of Distance from the Borderline. 'B' is the borderline condition where uncertainty is at the highest level; the numbers 1 through 6 indicate increasing distance from the borderline, with Step Size 6 representing an instance with the lowest uncertainty.

Figure 5. Mean response times from Experiment 2 for the Bargraph, Digital and Configural display groups as a function of Distance from the Borderline. 'B' is the borderline condition where uncertainty is at the highest level; the numbers 1 through 6 indicate increasing distance from the borderline, with Step Size 6 representing an instance with the lowest uncertainty.

Figure 6. Mean proportion correct in Experiment 2 for the Bargraph, Digital and Configural displays as a function of blocks of trials during Training and during the Extended Practice sessions.

Figure 7. Mean response times (in msecs) in Experiment 2 for the Bargraph, Digital and Configural displays as a function of blocks of trials during Training and during the Extended Practice sessions.

Fig. 1a. Configural Display

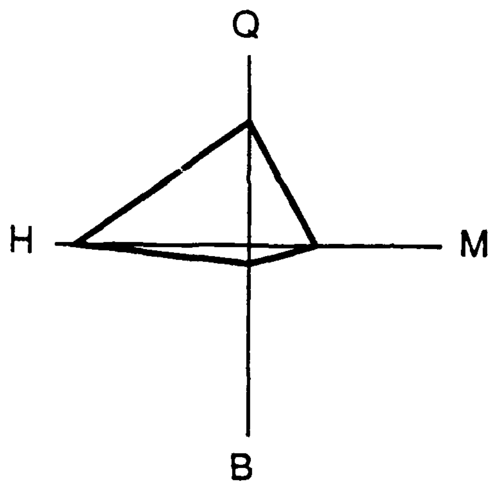

Fig. 1b. Digital Display

Q	M	B	H
66	**34**	**09**	**91**

Fig. 1c. Bargraph Display

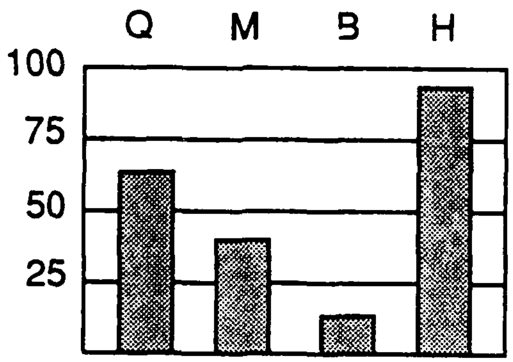

Fig. 2a: Training

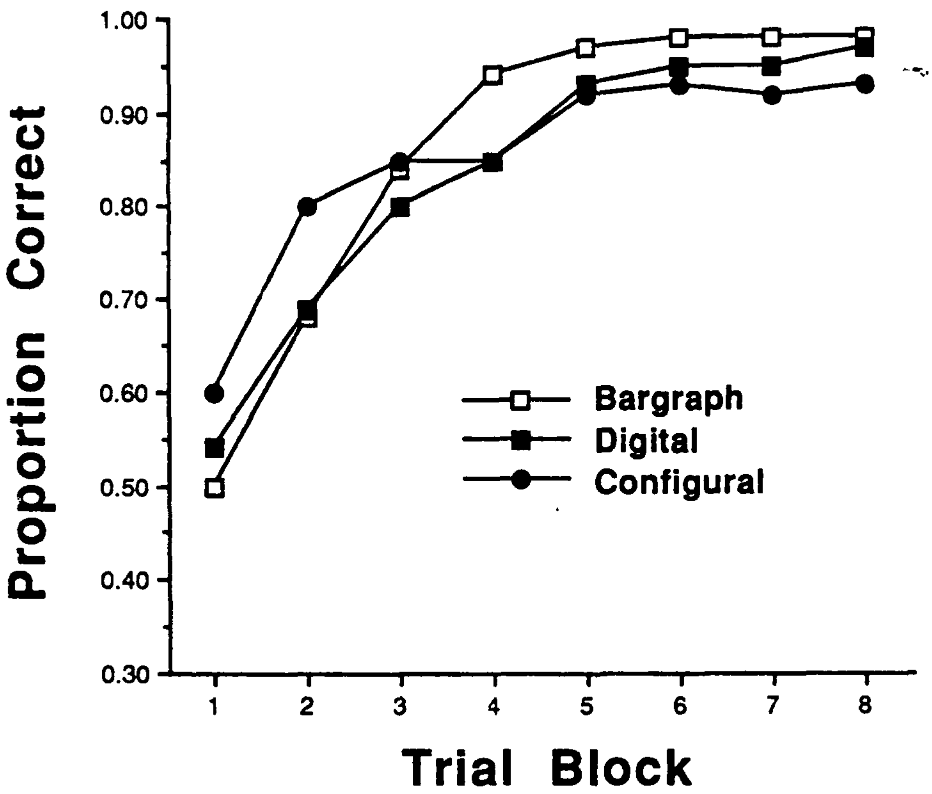

Fig. 2b: Extended Practice

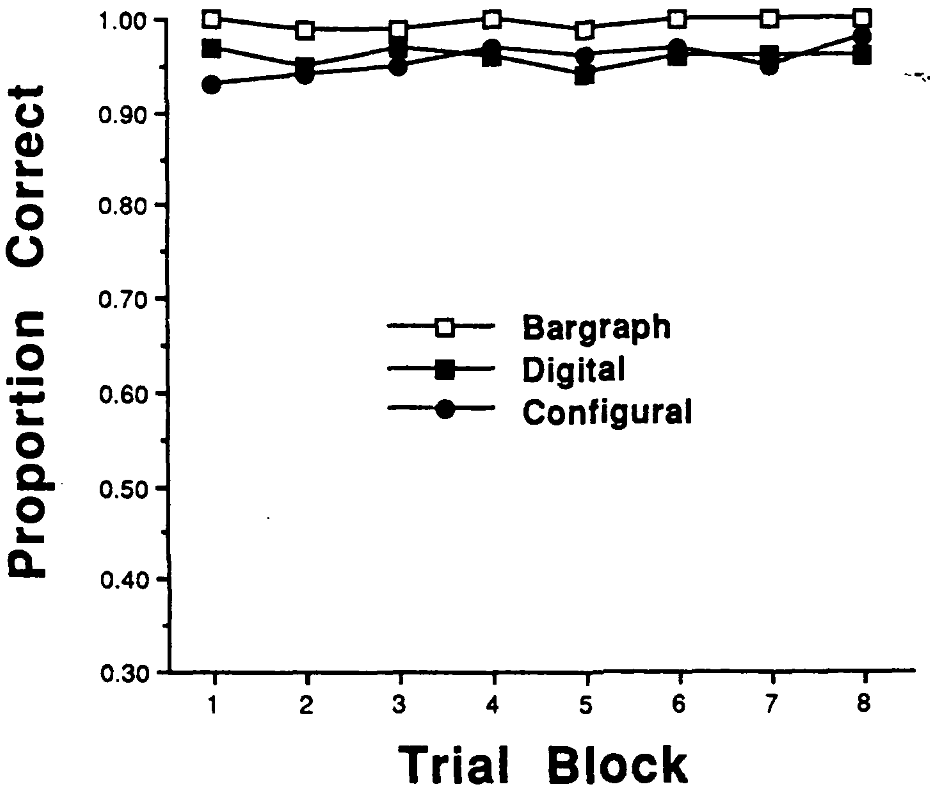

Fig. 3a: Training

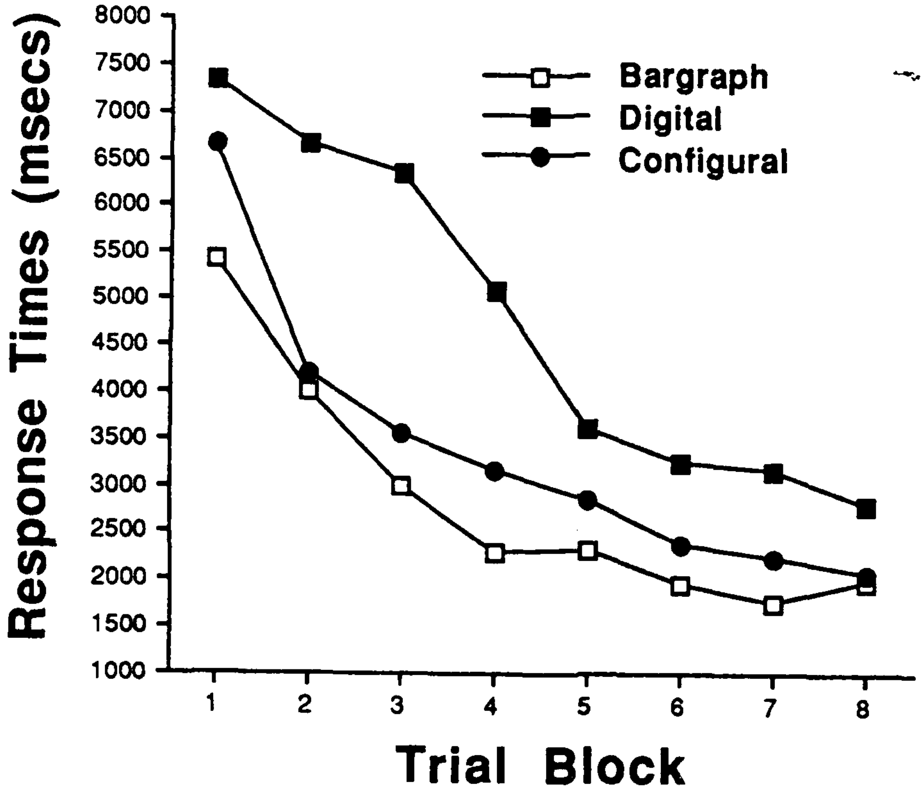

Fig. 3b: Extended Practice

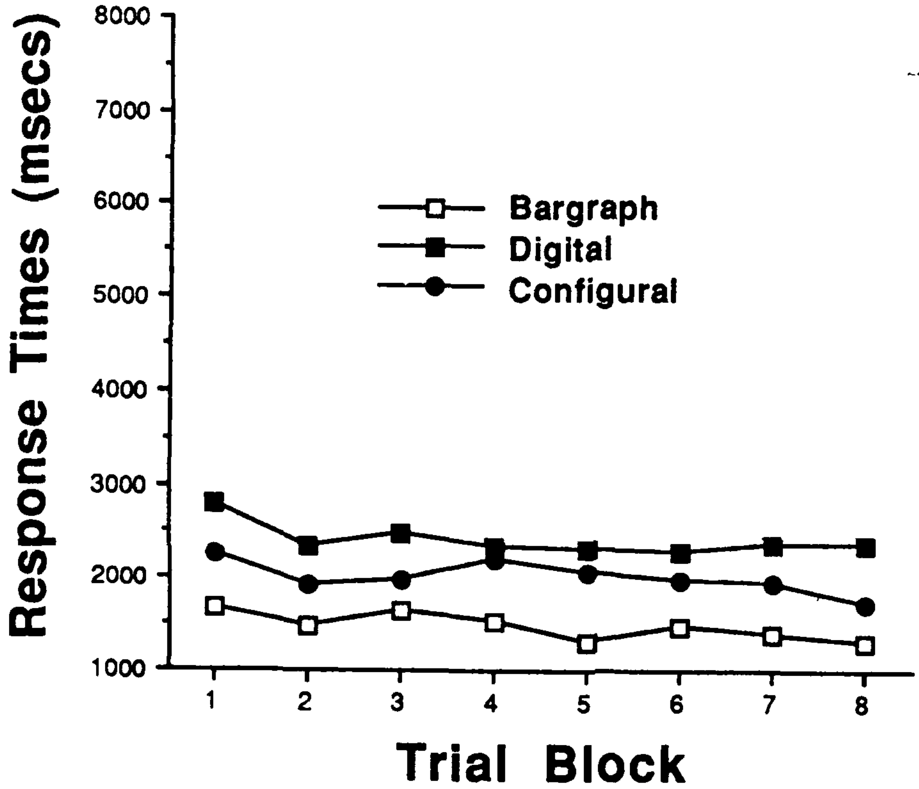

Fig. 4

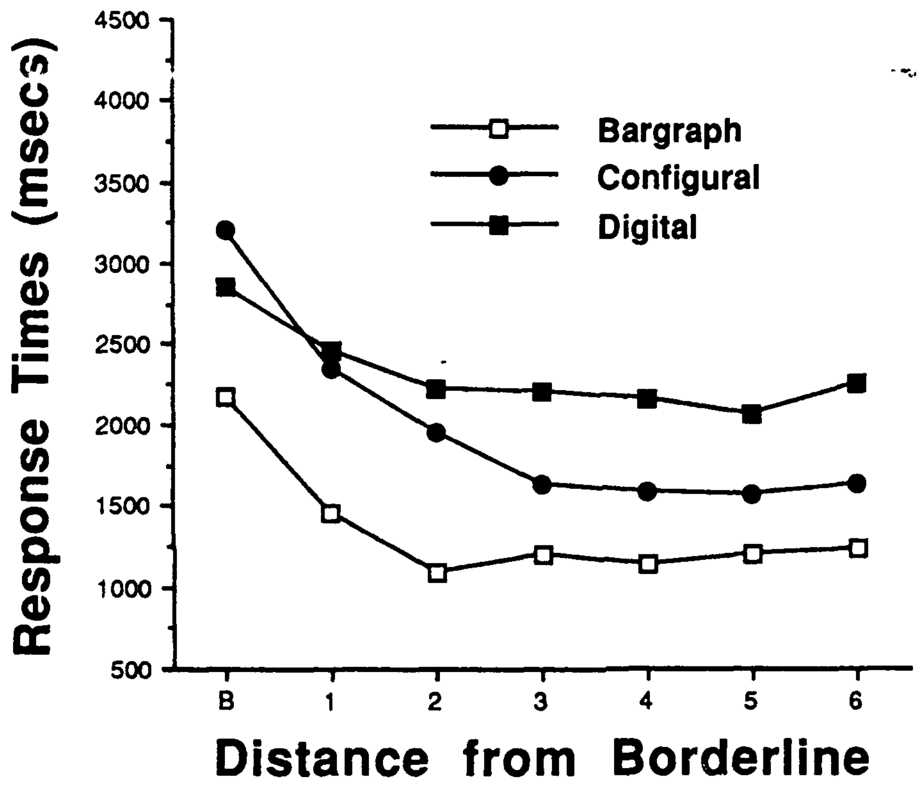
Response Times (msecs)
4500
4000
3500
3000
2500
2000
1500
1000
500
Bargraph
Configural
Digital
B
1
2
3
4
5
6
Distance from Borderline

Fig. 5

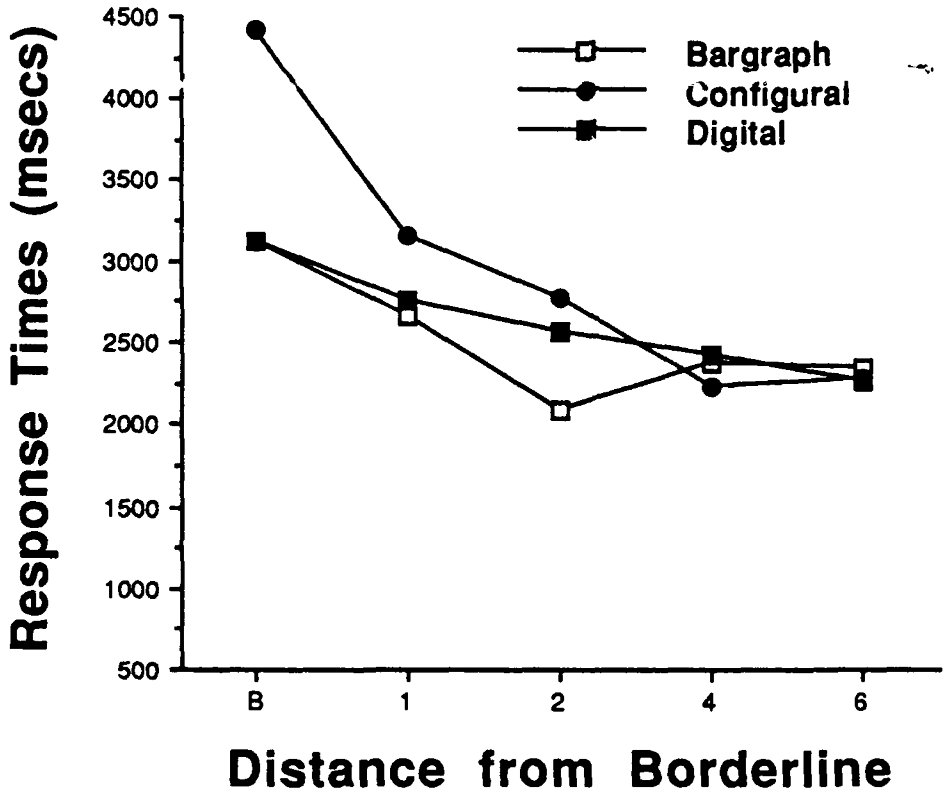
Response Times (msecs)
4500
4000
3500
3000
2500
2000
1500
1000
500
B
1
2
4
6
Distance from Borderline
Bargraph
Configural
Digital

Fig. 6a: Training

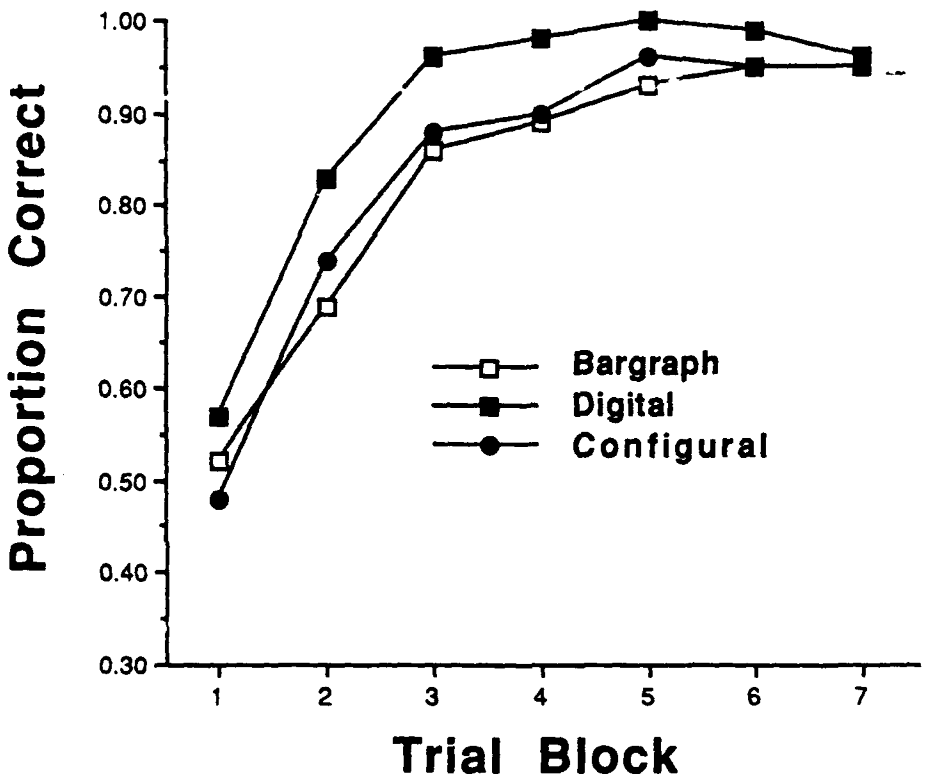

Fig. 6b: Extended Practice

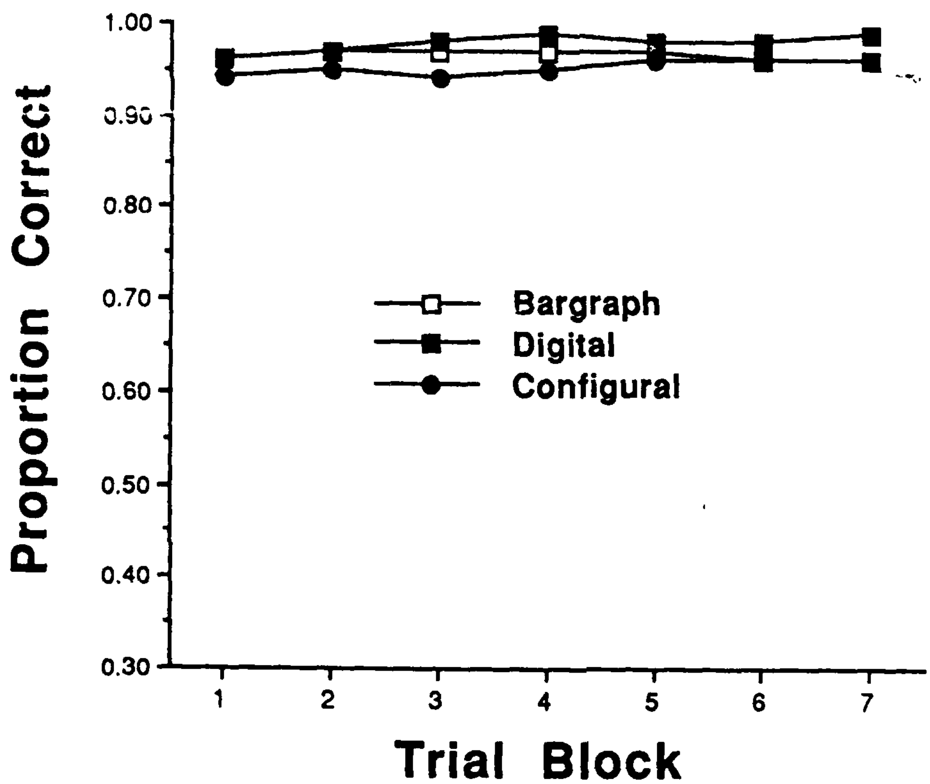

Fig. 7a: Training

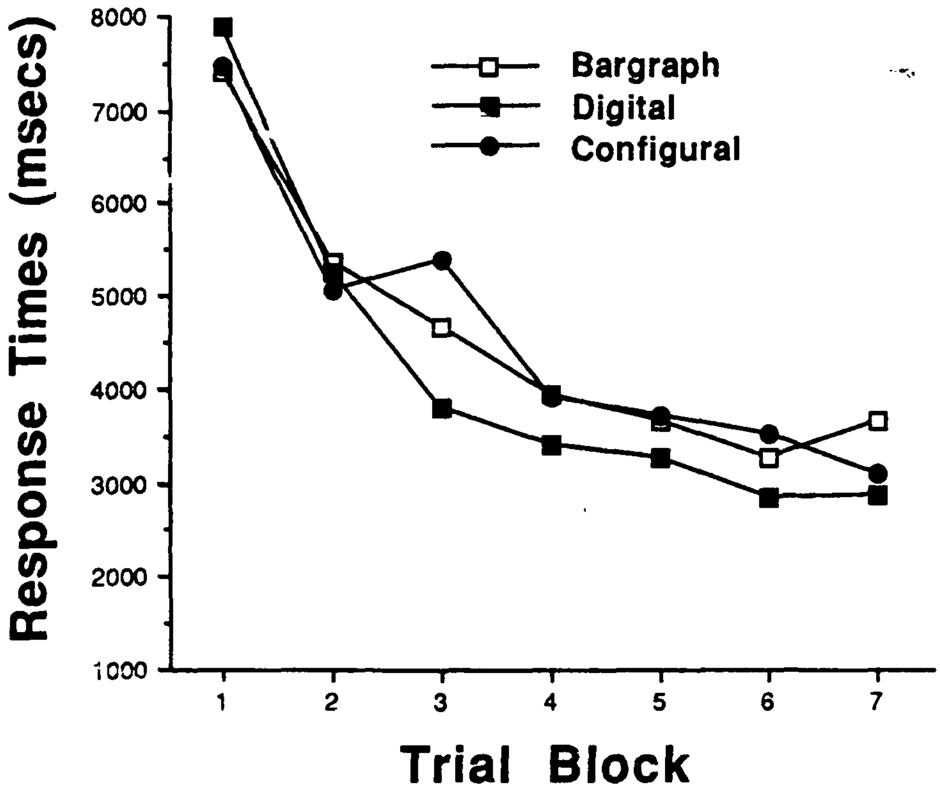

Fig. 7b: Extended Practice